CLAIMS

What is claimed is:

1.   A computer readable medium including executable instructions for processing training data for a statistical classification application, said computer readable medium comprising:

code for retrieving a plurality of training data structures that each comprise data members corresponding to feature elements and a data member identifying one of a plurality of classes;

code for processing each of said plurality of training data structures using probabilistic models that are a function of said feature elements to calculate a respective probability indicative of the respective training data structure belonging to its identified class; and

code for generating a scatter plot, using said plurality of training data structures, that visually indicates probabilities of said training data structures belonging to identified classes.

2.   The computer readable medium of claim 1 further comprising:

code for annotating points in said scatter plot to indicate probabilities of said plurality of training data structures belonging to identified classes.

3.   The computer readable medium of claim 1, wherein said code for generating a scatter plot displays points in said scatter plot using a predetermined color to indicate training data structures having probabilities below a threshold value.

4.   The computer readable medium of claim 1 further comprising:

code for identifying regions of said scatter plot that correspond to said plurality of classes.

5.   The computer readable medium of claim 1 further comprising:

code for receiving first input from a user to select a point corresponding to a respective training data structure.

6.   The computer readable medium of claim 5 further comprising:

code for displaying values of feature elements of said respective training data structure corresponding to said selected point.

7.   The computer readable medium of claim 5 further comprising:

code for displaying an image file associated with an object from which feature elements were derived in response to said code for receiving first input.

8.   The computer readable medium of claim 5 further comprising:

code for receiving second input from said user to reclassify said respective training data structure corresponding to said selected point.

9.   The computer readable medium of claim 8 further comprising:

code for revising said probabilistic models in response to said code for receiving said second input, wherein said code for processing is operable to recalculate probabilities of said plurality of training data structures belonging to identified classes using said revised probabilistic models.

10.   The computer readable medium of claim 5 further comprising:

code for receiving second input from said user to delete said respective training data structure corresponding to said selected point.

11.   A method for processing training data for a statistical classification application, the method comprising:

accessing a plurality of training data structures wherein each training data structure includes a plurality of feature variables and a variable identifying one of a plurality of classes;

calculating a respective confidence value for each of said plurality of training data structures that is indicative of a probability of the respective training data structure belonging to its identified class; and

generating a graphical user interface for a scatter plot that visually indicates confidence values for said plurality of training data structures.

12.   The method of claim 11 further comprising:

annotating at least a subset of points in said scatter plot with said confidence values.

13.   The method of claim 11 wherein said displaying uses a predetermined color to identify training data structures associated with a confidence value below a threshold value.

14.   The method of claim 13 wherein said threshold value is determined by receiving input from a user.

15.   The method of claim 13 wherein said graphical user interface identifies regions of said scatter plot associated with each of said plurality of classes.

16.   The method of claim 13 further comprising:

receiving user input to select a point of said scatter plot.

17.   The method of claim 16 further comprising:

displaying values of feature element variables of a training data structure corresponding to said selected point.

18.   The method of claim 16 further comprising:

displaying an image file associated with an object from which values, of a plurality of feature variables corresponding to said selected point, were obtained.

19.   The method of claim 16 further comprising:

deleting said training data structure corresponding to said selected point in response to further user input.

20.   The method of claim 16 further comprising:

reclassifying said training data structure corresponding to said selected point in response to further user input.

21.   The method of claim 11 further comprising:

refining probabilistic models after reclassification of at least one of said plurality of training data structures by a user; and

repeating said calculating and displaying in response to said refining.

22.   A system for processing training data for a statistical classification application, the system comprising:

means for processing a plurality of training data structures to generate a plurality of confidence values, wherein said each of said plurality of training data structures defines feature values and identifies one of a plurality of classes, wherein said confidence values indicate probabilities of objects having said feature values belonging to said identified classes; and

means for displaying a scatter plot using said plurality of training data structures that provides visual indication of probabilities of points belonging to identified classes.

23.   The system of claim 22 further comprising:

means for annotating points in said scatter plot to indicate probabilities of said plurality of training data structure belonging to identified classes.

24.   The system of claim 22 further comprising:

means for receiving first user input to select a point in said scatter plot.

25.   The system of claim 24 further comprising:

means for receiving second user input to reclassify a training data structure corresponding to said selected scatter point.

26.   The system of claim 25 further comprising:

means for revising probabilistic models associated with said plurality of classes, wherein said means for processing reprocesses said plurality of training data structures in response to said means for revising and said means for displaying redisplays said scatter plot using revised probabilities from said means for processing.